DOCUMENT RESUME

ED 236 172 TM 830 666

AUTHOR Mishra, Shitala P.
TITLE Ethnic Group Bias in WISC-R Verbal Items.
PUB DATE Aug 83
NOTE 19p.; Paper presented at the Annual Meeting of the American Psychological Association (Anaheim, CA, August 26-30, 1983). Light type throughout document.
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Ethnic Bias; *Intelligence Tests; Intermediate Grades; Maximum Likelihood Statistics; *Mexican Americans; Minority Groups; *Navajo; Standardized Tests; *Test Bias; Testing Problems; *Test Items; Verbal Tests
IDENTIFIERS Chi Square Analysis; *Wechsler Intelligence Scale for Children (Revised)

ABSTRACT

The study investigated cultural bias in the verbal items of the Wechsler Intelligence Scale for Children-Revised (WISC-R). Six verbal subtests of the WISC-R were administered to 40 Anglo, 40 Mexican-American, and 40 Navajo intermediate grade students matched for grade level, sex, and socioeconomic status. The responses of three groups of subjects to individual items comprising verbal subtests were analyzed by log-linear technique using the likelihood ratio chi-square statistic. The findings revealed that the performance of subjects was homogeneous across three cultural groups on the items comprising Arithmetic, Comprehension, and Digit Span subtests. Ten of the 79 items comprising Information, Similarities, and Vocabulary subtests were, however, found to be biased against Mexican-American subjects whereas 15 items of the same three subtests were identified as biased against Native-American Navajos. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED236172

Ethnic Group Bias in WISC-R Verbal Items

Shitala P. Mishra[1]

The University of Arizona

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

S. P. Mishra

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 830666

1. Paper presented at the annual meeting of the American Psychological Association, Anaheim, CA., August, 1983.

Abstract

The study investigated cultural bias in the verbal items of the Wechsler Intelligence Scale for Children-Revised (WISC-R). Six verbal subtests of the WISC-R were administered to 40 Anglo, 40 Mexican-American, and 40 Navajo subjects matched for grade level, sex, and socioeconomic status. The responses of three groups of subjects to individual items comprising verbal subtests were analyzed by log-linear technique using the likelihood ratio chi-square statistic. The findings revealed that the performance of subjects was homogeneous across three cultural groups on the items comprising Arithmetic, Comprehension, and Digit Span subtests. Ten of the 79 items comprising Information, Similarities, and Vocabulary subtests were, however, found to be biased against Mexican-American subjects whereas 15 items of the same three subtests were identified as biased against Native-American Navajos.

In recent years there have been a number of research attempts to examine and identify fundamental issues concerning test bias and assessment of minorities (Anastasi, 1967; Oakland, 1977). Historically such attempts, however, reflect a heavy concentration on the issue of predictive validity (Shepard, 1982). Such an approach to predictive validity focuses on the relationship between test scores and criterion measures such as academic success or job performance. A psychometric instrument, according to predictive validity approach, is considered to be fair if the regression equation for one cultural group does not over-or-under predict performance on the common criterion for other ethnic populations (Cleary, 1968).

Although fairness of a test in predicting external criteria is important for the purposes of ensuring equal opportunity in selection and placement, the adequacy of the predictive validity approach to the problem of test fairness has been questioned by Thorndike (1971). Additionally, evidence of predictive validity does not preclude the presence of item bias in a test and yet it is at the level of individual items that serious charges of test bias have been made (Larry P. vs. Wilson Riles, Note 1; Shepard, 1982). Many critics of mental tests such as Williams (1971) have suggested that certain items on currently used standardized

tests are likely to be more difficult for blacks since opportunity to learn the content of such items is less available to these children in their cultural environment.

Despite increasing need for more data on item bias provide direct and basic information for answering popular charges of test bias, a surprisingly limited number of empirical studies have explored the differential effects of culture upon items of individually administered tests. The findings of item bias studies may be especially important in determining the relationship between an item and a particular construct, such as intelligence, for various cultural groups. If the validity of such relationships are not generalizable across different ethnic groups, then the appropriateness of of item content with populations of ethnic minorities becomes questionable. The information concerning item bias is not only useful in addressing the issue of fairness of test content; it can also assist practicing psychologists in interpreting the meaning of the performance of an individual child across different types of test items.

Although an increasing number of studies in recent years have examined item bias with standardized tests, the variety of approaches and conclusions emerging from research findings still seem unable to generaate consumer confidence in the use of tests like the WISC-R. Since the use of ability tests in clinical and diagnostic settings goes beyond consideration of predictive validity a careful scrutiny of items believed to

represent a common construct across ethnic groups is important. The purpose of the present study, therefore, was to report additional data on the ethnic group bias in the verbal items of the WISC-R. Since the WISC-R is the most frequently used test for psychoeducational assessment of children in educational as well as clinical settings, the findings of this study may provide some assistance to researchers and practitioners in the assessment of minority subjects.

Method

Sample

The subjects for this study were selected from fourth and fifth grade students attending elementary schools. The populations were first stratified in terms of ethnicity, grade level, sex, and socioeconomic status, and then a table of random numbers was used to select 40 Anglo, 40 Mexican-American, and 40 Navajo subjects. All samples included approximately equal number of male and female subjects. In the process of sampling, an attempt was made to ensure the equivalency of SES across groups. Most of the subjects from three ethnic populations came from lower- and middle-class homes.

Testing Procedure

All subjects were individually administered the 10 regular subtests of the WISC-R by two certified school psychologists who had extensive experience in administering and scoring individual as well as group tests for subjects of varying ages and cultures. The standard procedures outlined

in the test manual were followed closely. Even though all regular subtests of the WISC-R were administered to all subjects, only six verbal subtests were used in the present study as items on these subtests have been found to be somewhat more sensitive to cultural bias in Mexican-American groups (Cotter & Berk, Note 2).

Data Analysis

The obtained data were analyzed using the log-linear model developed by Goodman (1978) for analyzing multi-dimensional categorical data. The log-linear models are based on maximum likelihood methods and permit incorporation of all interactions in the multidimensional contingency table without the need to designate some of the variables as dependent on others. Such an analysis made it possible to test the interaction effects among ethnicity, ability, and pass-fail response categories. By taking the natural logarithm (log) of the respective values, the log-linear model transforms the multiplicative relationship between expected cell frequencies and marginal totals to an additive relationship; this makes it possible to analyze interactions similar to those performed on interval data (Everitt, 1977; Fienberg, 1979). Goodness-of-fit is tested with the likelihood chi-square statistic (L^2) (Everitt, 1977). Under the null hypothesis, L^2 is distributed like X^2 and has the same degrees of freedom. In multidimensional analysis, L^2 has the advantage of being partitioned exactly when interactions are tested while the Pearson X^2 is only approximate.

The procedure for evaluating the fit of models to the data in log-linear analyses involves a high probability of Type II error. Power can be increased by incresing the probability of Type I error, and it is argued that traditional alpha levels of .01 or .05 may be unnecessarily stringent protection against rejecting the null hypothesis when it is true (Knoke & Burke, 1980). A detailed description and rationale for the use of log-linear procedures are beyond the purpose and scope of this study but can be found in Everitt (1977) and Goodman (1978). The BMDPSF computer program (Dixon, 1979) was used to perform the L^2 analysis.

The three variables considered in the study were ethnicity, ability, and pass-fail responses to individual items. The three ethnicity categories were Anglo, Mexican-American, and Navajo subjects. The two ability categories were created within each group on the basis of subjects' Verbal IQs (VIQ) on the WISC-R. The VIQs in the low ability category for three ethnic groups ranged from a low of 72 to a high of 98 while the high ability subjects in all groups represented VIQs from 98-123. Using total scores on the same test for creating homogeneous intervals has been recommended for controlling the effects of ability differences between subpopulations being studied (Sheuneman, 1980).

Pass-fail responses to individual items on six selected subtests served as two levels of the third dimension going into three dimensional contingency table analysis. This analysis permitted testing various models ranging from

complete independence of all the variables to the effects of combined variables on item scores. However, the major concern of the study was to examine the ethnic bias in item performance by controlling the effects of ability. Consequently, the model pertaining to independence of ethnicity and item performance while holding ability constant was tested for all the items on the six verbal subtests of the WISC-R.

Results

Scores obtained by 40 Anglo and 40 Mexican-American subjects were analyzed item by item for each of the six verbal subtests of the WISC-R. According to the scoring criteria outlined in the WISC-R manual (Wechsler, 1974) some of the items on these subtests are scored as 0 (fail) or 1 (pass) whereas other items are assigned scores ranging from 0 to 2. Exact item scores, however, were not examined in this study, instead a score of 0 was considered a failure while any score value above 0 was considered a passing score and was coded as 1. Thus, all the responses were scored as 0 (fail) or 1 (pass) to provide a consistent dichotomous measure for the analysis.

Comparison of Anglo and Mexican-American Subjects

An examination of the frequency of correct responses by subjects from Anglo and Mexican-American ethnic groups revealed no significant differences in the percentage of passing the items comprising the Comprehension, Digit Span, and the Arithmetic subtests. Only 10 items from the

Information, Similarities, and the Vocabulary subtests were identified as biased and are displayed in Table 1. It is evident that 27 out of 30 items on the Information subtest were passed or failed by an approximately equal number of subjects from each ethnic group, suggesting no significant differences in obtained frequency of subjects from each group passing or failing these items. For the remaining three items of this subtest, a significant difference between the percentage of subjects passing these items in the two ethnic groups was observed. The items detected to be biased against Mexican-American subjects were the twelfth, fifteenth, and eighteenth items on the Information subtest.

Insert Table 1 about here

Similarly, 2 out of 17 items on the Similarities subtest reflected evidence of bias against Mexican-American subjects. On 14 items subjects from both ethnic groups showed no significant differences in performance. On the eighth and ninth items of this subtest, a significantly lower percentage of Mexican-American subjects passed these items as compared to their Anglo counterparts.

The performance of two groups of subjects on Vocabulary subtest was found to be homogeneous for 25 out of 30 items. This indicated that approximately 83% of the items on this test seemed to be fair for Mexican-American subjects. There were five items which did, however, provide an indication of

a significant discrepancy in the percentage of passing these items by subjects from two ethnocultures. The systematic difference in performance was observed for items 8, 9, 13, 16, and 19, suggesting significantly lower percentage of Mexican-American subjects passing these items. Considering all of these three subtests of the WISC-R, approximately 13 percent of the 79 items seemed to be biased or more difficult for Mexican-American subjects.

Comparison of Anglo and Navajo Subjects

The frequency of correct responses by subjects from Anglo and Navajo groups on 15 items identified as biased is displayed in Table 2. An examination of the findings revealed that only 15 items from the Information, Similarities, and the Vocabulary subtests were identified as biased against Navajo sample. However, the performance of Anglo and Navajo subjects revealed no significant differences across all items of the Completion, Digit Span, and the Arithmetic subtests.

Insert Table 2 about here

For the five items of the Information subtest, a significant difference in the percentage of subjects passing the items in the two ethnic groups was observed. The items detected to be biased were the twelfth, fifteenth, sixteenth, seventeenth, and nineteenth items on the Information subtest.

The obtained L^2 values for the items on the Similarities subtest suggested that 4 out of 17 items of this test reflected bias against Navajo subjects. On 13 items, subjects from both ethnic groups did not show any significant differences in their performance. On the tenth, eleventh, thirteenth, and fourteenth items of the Similarities subtest,a significantly lower percentage of Navajo subjects passed these items as compared to Anglo subjects. The L^2 for these four items ranged from 6.31 to 11.32, and was found to be significant at the .05 or lower level of confidence.

The performance of the two groups of subjects was found to be similar for 24 out of 30 itms on the Vocabulary subtest. This indicated that 80% of the items on this test seemed to be fair for Navajo subjects. There were six items, however, that provided an indication of significant discrepancy in the persentages of subjects passding in the two ethnic groups. Such differences in performance were noticed for items 9, 16, 17, 18, 19, and 20. The obtained values of L^2 ranged from 6.01 to 27.47, and were found to be significant at the .05 or lower level of confidence. Considering items on all three selected verbal subtests of the WISC-R, it was observed that approximately 19% of the 79 items seemed to be biased or more difficult for the Navajo subjects, as indicated by the lower percentage of Navajo subjects passing these items, compared to their Anglo counterparts.

Discussion

The performance of Anglo, Mexican-American, and Navjo subjects on verbal items of the WISC-R clearly suggested that the majority of the items are free from cultural bias. Only 15 items provided an indication of non-homogeneous performance by subjects from three different cultural backgrounds. These items are from the Information, Similarities, the Vocabulary subtests. Such a discrepancy in performance of three groups of subjects on some of the items of three subtests detected to be biased in this study are consistent with the findings of recent investigations in this area. In the black-white comparisions, for example, item-by-group interactions were reported by Cotter and Berk (Note 2) for the Information and Similarities subtests. Item 12 on the Information subtests has also been used to present the evidence of bias against blacks in the most publicized Larry P. vs. Wilson Riles court case (Note 1).

Some of the Vocabulary items identified as biased in the present study were also reported being as culturally sensitive by Martin (Note 5) using a Cuban-Spanish translation of the WISC-R. Martin (Note 5) found that rank order of item difficulty for some Vocabulary items changed for his Cuban sample as compared to rank order of item difficulty obtained for the normative sample. Findings of some of these recent investigations in this area may, therefore, lend some support to the notion that the experiential backgrounds of culturally different subjects may not have provided opportunities to

learn the content of certain vocabulary items (Murray, Note 3; Samuda, 1975) contrary to the notion that the vocabulary items in the WISC-R allow "any recognized meaning of the word disregarding any elegance of expression" (Wechsler, 1974, p. 161). It should be noted, however, that significant differences in the percentage of individuals passing or failing an item were evident for only a very few verbal items. Since a majority of the items were found to be free from cultural bias, the overall direction of these findings seems to be consistent with the evidence provided by recent studies indicating absence of item bias in the WISC-R (Sandoval, 1979). However, somewhat contradictory findings with respect to ten items of the Information, Similarities, and Vocabulary subtests should be interpreted with appropriate caution as limited sample size may restrict the generalizability of these findings. It should also be noted that the investigation of the item bias issue within the context of this study involved an examination of the probability of responding to an item correctly for subjects of equal ability regardless of their ethnic group membership. It was, therefore, necessary to score all item responses on a pass-fail basis. Use of such a dichotomous scoring procedure did not permit the examination of possible bias in 1 versus 2 point responses. However, in considering conclusions of item bias it might be of some practical significance for future studies to also examine the impact of culture on item responses that are assigned multiple score values.

Reference Notes

1. Larry P. vs. Wilson Riles, U.S. District Court for the Northern District of California, 1972, No. C-712270 RFP.

2. Cotter, D. E. & Berk, R. A. Item bias in the WISC-R using black, white, and hispanic learning disabled children. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, April, 1981.

3. Murray, A. M. Item bias in the McCarthy Scales of Children's Abilities for Anglo and Mexican-American Children. Unpublished doctoral dissertation, The University of Arizona, 1981.

4. Sandoval, J., Zimmerman, I. L., & Woo-Sam, J.M. Cultural differences on WISC-R verbal items. Paper presented at the annual meeting of the American Psychological Association, Montreal, September, 1980.

5. Martin, P. C. A Spanish translation, adaptation, and standardization of the Wechsler Intelligence Scale for Children-Revised. Unpublished doctoral dissertation, The University of Miami, 1977.

References

Anastasi, A. Psychological testing (5th ed.). New York: Macmillan, 1982.

Berk, R. A. Introduction. In. R. A. Berk (Ed.), Handbook of methods for detecting test bias. Baltimore: The Johns Hopkins University Press, 1982.

Cleary, T.A. Test bias: Prediction of grades of negro and white students in integrated colleges. Journal of Educational Measurement, 1968, 5, 115-124.

Dixon, W. J. BMDP-Biomedical Computer Programs, Berkeley: University of California Press, 1979.

Everitt, B. S. The analysis of contingency tables. New York: John Wiley & Sons, Inc., 1977.

Fienberg, S. E. The analysis of cross-classified categorical data. Cambridge, Mass.: MIT Press, 1977.

Goodman, L. A. Analyzing qualitative/categorical data. Cambridge, Mass.: ABT Books, 1978.

Knoke, D. & Burke, P. J. Log-linear models. Beverly Hills, CA: Sage Publications, 1980.

Miele, F. Cultural bias in the WISC. Intelligence, 1979, 3, 419-164.

Oakland, T. (Ed.). Psychological and educational assessment of minority children. New York: Brunner/Mazel, 1977.

Samuda, R.J. The psychological testing of American minorities. New York: Harper & Row, 1975.

Sandoval, J. The WISC-R and internal evidence of test bias with minority groups. Journal of Consulting and Clinical Psychology, 1979, 47, 919-927.

Scheuneman, J. D. Latent-trait theory and item bias. In L. J. Th. van der Kamp, W. F. Langerak, & D. N. M. de Gruizter (Eds.), Psychometrics for educational debates. New York: John Wiley & Sons, 1980.

Shepard, L. A. Definitions of bias. In R. A. Berk (Ed.), Handbook of methods for detecting test bias. Baltimore: The Johns Hopkins University Press, 1982.

Thorndike, R. L. Concept of culture-fairness. Journal of Educational Measurement, 1971, 8, 63-70.

Vance, H. B., Gaynor, P., & Coleman, M. Analysis of cognitive abilities for learning disabled children. Psychology in the Schools, 1976, 13, 477-483.

Wechsler, D. Wechsler Intelligence Scale for Children-Revised. New York: The Psychological Corporation, 1974.

Williams, R. L. Abuses and misuses in testing black children. Counseling Psychologist, 1971, 2, 62-73.

Table 1

Summary of L^2 Analysis for WISC-R Items Tested with the Log-Linear Procedure for Independence of Ethnicity and Item Performance

WISC-R Subtest	Item Number	Total Passing in Both Groups	Percent Passing				L^2
			Anglo		Mexican-American		
			N	%	N	%	
	12	38	30	78.94	8	21.06	14.47**
Informa-	15	9	9	100.00	0	00.00	10.93**
tion	18	24	20	83.33	4	16.67	5.76*
Similari-	8	48	34	70.84	14	29.16	6.96*
ties	9	42	31	73.81	11	26.19	6.53*
	8	58	39	67.25	19	32.75	5.38*
	9	51	36	70.59	15	29.41	5.44*
Vocabulary	13	34	26	76.48	8	23.52	6.31*
	16	30	29	94.34	5	16.67	9.96**
	19	13	13	100.00	0	00.00	8.67**

*p < .05
**p < .01

Note 1: Obtained likelihood chi-square values were tested for significance with df=2.

2: % passing refers to the proportion of subjects passing an item in each group in relation to the total number of correct item responses from both groups combined.

Table 2

Summary of L^2 Analysis for 15 WISC-R Items Tested with the Log-Linear Procedure for Independence of Ethnicity and Item Performance

WISC-R Subtest	Item Number	% Passing Anglo	% Passing Navajo	L	L^2
Information	12	75.0	37.5		19.11
	15	22.5	5.0		10.63
	16	25.0	2.5		9.07
	17	35.0	2.5		20.07
	19	32.5	0.0		16.20
Similarities	10	32.5	5.0		11.31
	11	42.5	12.5		10.94
	13	27.5	0.0		6.57
	14	20.0	0.0		6.31
Vocabulary	9	90.0	72.5		5.83
	16	62.5	10.0		25.99
	17	45.0	0.0		27.47
	18	32.5	0.0		13.92
	19	50.0	10.0		12.67
	20	20.0	0.0		6.44

Note: Obtained likelihood chipsquare values were tested for significance with df = 2.